**COMPARISON OF MACHINE LEARNING ALGORITHMS FOR ACTIVITITY**

**CLASSIFICATION USING ELECTOMYOGRAPHY**

1,2 Anthony Anderson, 1,2 Andrea Willson, 1,2 Evan Schuster, and 1,2 Patrick Aubin, PhD

1 University of Washington, Seattle, WA, USA

2 VA Center for Limb Loss and Mobility, Seattle, WA, USA

email: paubin@uw.edu, web: https://www.amputation.research.va.gov/

# **INTRODUCTION**

Abundant wearable sensor data and advances in machine learning algorithms have made activity classification viable for biomechanical studies.

# **METHODS**

With approval from the VA Puget Sound Health Center and University of Washington IRBs, EMG data was recorded for two subjects (male, age 24-27, height 5’10”-6’2”) at the VA Center for Limb Loss and Mobility. For each subject, 16 EMG sensors (Delsys Inc., Natick MA) were placed on the right leg and lower abdomen. Before adhering the electrodes, the sensor locations were shaved and wiped with alcohol. Ten of the sensors were placed on primary leg muscles commonly measured in EMG studies. These muscle locations were found via palpations and references to anatomical landmarks. The six remaining sensors were placed pseudo-randomly, without isolating specific muscles (Table 1). All sensors wirelessly streamed data to a desktop PC at 1200 Hz for the duration of the study. For both subjects, 150 three-second trials were recorded while the subject performed each of the following activities: standing, sitting, walking at a self-selected pace, walking at 175% of the self-selected pace, stair ascent, and stair descent.

During data analysis in MATLAB (MathWorks, Natick MA), each three-second recording was separated into 0.15-second windows. Each window contained signals from all 16 sensors and was labeled as one of the six activities. The following five features were then extracted for each sensor over each of the 0.15-second snapshots: mean average value, number of zero crossings, variance, number of slope sign changes, and waveform length (sum of the vertical point-to-point distance over the window). These fives features across sixteen sensors resulted in an 80-dimensional feature vector for each window.

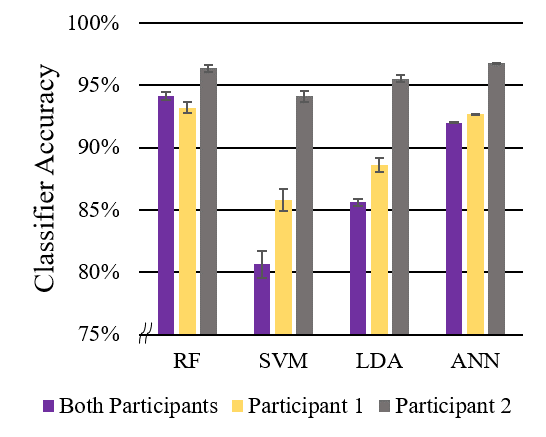
The data then was randomly split into a training set (80% of data) and a testing set (20%) of data. The training data was reduced via SVD to a collection of 40-dimensional feature vectors that retained 99% of the variance in the initial feature set. These labeled vectors were then used to train the following classification algorithms: linear discriminant analysis (LDA), a random forest (RF) with 40 decision trees, a linear support vector machine (SVM), and an artificial neural network (ANN) with 10 hidden layers. Each classifier was then tested on its ability to accurately predict the activities in the testing set.

This train-test process was repeated ten times with a different random 80-20 cut of the initial data set each time. Additionally, the classifiers were trained on an individual basis, i.e. trained with Participant 1’s data and tested with Participant 1’s data, as well as on a group basis where the algorithms were both trained and tested with both participants’ data.

To determine how classification accuracy was impacted by removing sensors from the data, a backward-selection algorithm was implemented for each classifier. All classifiers were trained fifteen times, every time with a different sensor removed from the training and testing sets. The sensor that reduced the classifier accuracy by the least amount was then permanently removed from the set. This process was repeated until only one sensor remained. Feature vectors were cast into SVD bases but were not truncated as before, as the size of the feature vector changed as sensors were removed. This resulted in higher classification accuracies but substantially longer training times.

# **RESULTS AND DISCUSSION**

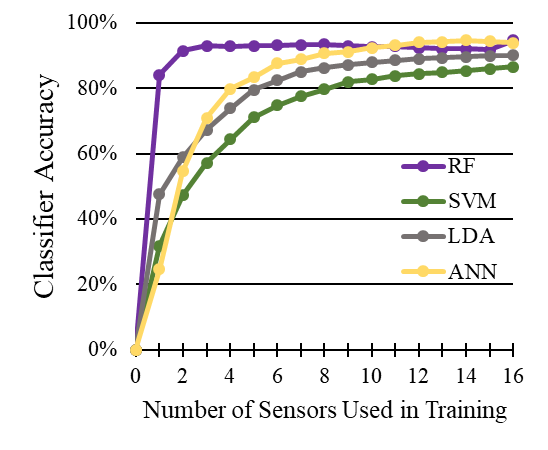
All the classifiers were able to predict which activity produced a given feature vector with high accuracy (Figure 1). The classifier with the highest accuracy when being trained and tested with data from both participants was the RF at 94.17 ± 0.32%. The classifier with the lowest accuracy was the SVM at 81.63 ± 0.11%.



**Figure 1:** Mean classifier accuracies when trained on individual and combined subject data with all sensors.

While the accuracy of the RF and ANN were comparable, the backward-selection algorithm showed the RF to be superior, as it could achieve over 90% accuracy with only two sensors (Figure 2). The ANN required eight sensors to reach a comparable classification accuracy.

The backward-selection algorithm also revealed that the classifiers did not rely exclusively on the sensors placed on muscle bellies but used information from the nonconventional sensor placements as well.



**Figure 2:** Classification accuracy as a function of number of sensors used to train the classifier.

These results can be applied in the context of prosthetic limb control or monitoring patient activity outside of the laboratory environment.

# **REFERENCES**

1. Sharafi and Blemker. *Proceedings of NACOB’08,* Ann Arbor, MI, USA, 2008.
2. XXX

**Table 1:** Sensor locations. Sensors with yellow shading were placed without regard to muscle location.

